nature COMMUNICATIONS

ARTICLE

Received 6 May 2014 | Accepted 7 Nov 2014 | Published 17 Dec 2014

DOI: 10.1038/ncomms6803

Hydrogen bond rotations as a uniform structural tool for analyzing protein architecture

Robert C. Penner[1,2], Ebbe S. Andersen[3,4,5], Jens L. Jensen[6], Adriana K. Kantcheva[4,7], Maike Bublitz[4,7], Poul Nissen[3,4,7], Anton M.H. Rasmussen[3,8], Katrine L. Svane[3,8], Bjørk Hammer[3,8], Reza Rezazadegan[1], Niels Chr. Nielsen[3,9,10], Jakob T. Nielsen[3,9,10] & Jørgen E. Andersen[1,2,6]

Proteins fold into three-dimensional structures, which determine their diverse functions. The conformation of the backbone of each structure is locally at each C^{α} effectively described by conformational angles resulting in Ramachandran plots. These, however, do not describe the conformations around hydrogen bonds, which can be non-local along the backbone and are of major importance for protein structure. Here, we introduce the spatial rotation between hydrogen bonded peptide planes as a new descriptor for protein structure locally around a hydrogen bond. Strikingly, this rotational descriptor sampled over high-quality structures from the protein data base (PDB) concentrates into 30 localized clusters, some of which correlate to the common secondary structures and others to more special motifs, yet generally providing a unifying systematic classification of local structure around protein hydrogen bonds. It further provides a uniform vocabulary for comparison of protein structure near hydrogen bonds even between bonds in different proteins without alignment.

[1] Centre for Quantum Geometry of Moduli Spaces, Aarhus University, DK-8000 Aarhus C, Denmark. [2] Departments of Mathematics and Theoretical Physics, Caltech, Pasadena, California 91125, USA. [3] Interdisciplinary Nanoscience Center, Aarhus University, DK-8000 Aarhus C, Denmark. [4] Department of Molecular Biology and Genetics, Aarhus University, DK-8000 Aarhus C, Denmark. [5] Centre for DNA nanotechnology, Aarhus University, DK-8000 Aarhus C, Denmark. [6] Department of Mathematics, Aarhus University, DK-8000 Aarhus C, Denmark. [7] Centre for Membrane Pumps in Cells and Disease, Aarhus University, DK-8000 Aarhus C, Denmark. [8] Department of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark. [9] Department of Chemistry, Aarhus University, DK-8000 Aarhus C, Denmark. [10] Center for Insoluble Protein Structures, Aarhus University, DK-8000 Aarhus C, Denmark. Correspondence and requests for materials should be addressed to J.E.A. (email: andersen@qgm.au.dk).

A hydrogen bond[1] forms between an electronegative atom (the acceptor) and a hydrogen atom covalently bound to another electronegative atom (the donor). Hydrogen bonds are of key importance in determining and fine-tuning molecular structure, interaction, function[2] and specificity of molecular recognition[3]. It is widely recognized that the function of a protein is intimately linked to the three-dimensional (3D) structure of its native folded state. Besides polypeptide assembly through covalent bonds, the structure is determined and stabilized by van der Waals interactions, hydrophobic packing, hydrogen bonds and ionic interactions. As hydrogen bonds are readily broken and reformed, they determine alternative conformations, and hence are also important for conformational changes of proteins.

Dihedral angles specify the backbone conformation of proteins by providing a complete 2D description of the progression from one peptide unit to the next along the backbone, as displayed in the well-known Ramachandran plots[4]. However, dihedral angles are not reliable when considering relative configurations, which are separated far apart along the backbone.

As the geometric phase space of hydrogen bonds has large dimension *a priori*, 3D and 4D simplifications have captured only part of their geometry[5,6]. Here, we introduce a systematic 3D descriptor of main chain hydrogen bond geometry by assigning to each hydrogen bond between backbone C = O and N–H atoms a spatial rotation, which is evidently independent of the overall spatial orientation of the protein just as for dihedral angles. The rotation is simply obtained as follows: First, rotate the entire protein, such that the donor peptide unit is aligned with the standard coordinate axes; the subsequent rotation from this standard location of the donor to the acceptor is the 3D descriptor of the hydrogen bond. As any rotation is given by a rotation axis and degree of rotation, it can be plotted in 3D space as a vector along the rotation axis of length equal to the degree of rotation. Hence all vectors in 3D space of length at most π describe all rotations. Rotations can thus be displayed by semi-transparent 2D projections on the page to graphically represent the geometry of main chain hydrogen bonds, akin to Ramachandran plots of dihedral angles for backbone geometry.

Using a set of high-quality structures culled from the PDB[7] to probe the conformational space, we measure the rotations for the totality of hydrogen bonds (1.16 M hydrogen bonds) from all these structures and find, amazingly, that all the rotations concentrate into 30 well-defined clusters, which together comprise just over 32% of the volume of rotational space. In fact, 93.4% of all the rotations are contained in the seven biggest clusters occupying just under 20% of the volume of rotational space. This is particularly striking, as we find that fully 95% of the volume of rotational space is accessible when describing atoms as rigid spheres, whereas still 50% is available when modelling the hydrogen bonding of peptides with Density Functional Theory (DFT), accepting all bonds of at least 0.1 eV hydrogen bond strength. By analyzing the resulting data, we conclude that this rotation descriptor alone determines the relative configurations of the two peptide planes involved in the hydrogen bond, with one exception, which is explained in detail below.

Our clusters correlate with well-known patterns such as α, 3_{10} and π-helices, parallel and antiparallel *β*-structures, and further provide a collection of motifs found in random coil and turn elements of protein structure. All of this is obtained from our new systematic uniform viewpoint of rotations associated to backbone hydrogen bonds.

In this way, the rotational descriptor provides a uniform viewpoint on local structural motifs around main chain hydrogen bonds in proteins. Practically speaking, as each such cluster displays specific structural characteristics, the associated classification of hydrogen bond geometry is useful in studying specific protein function and conformational changes. It further gives a novel vocabulary and a quantitative measure by which one can compare local configurations around hydrogen bonds within each structural element of a protein as well as between such, and even for hydrogen bonds from different proteins without requiring alignment.

Results

The rotational descriptor of main chain hydrogen bonds. We associate a triple of orthogonal unit vectors (that originated in earlier work[8]) to the *i*th peptide unit P_i along the backbone by using only main chain coordinates. The first vector is the unit vector from the centre of the carbon atom C to the centre of the nitrogen atom N in the *i*th peptide bond. The second vector is obtained by rotating the first vector 90° towards the oxygen atom O in the same peptide plane, and the last vector is the cross-product of the two first (Fig. 1a and Methods). For each *i*, there is a unique rotation $\mathcal{R}_{P_i}$, which brings the unit vectors parallel to the *x*, *y* and *z* axes to the triple associated to the *i*th peptide unit P_i. To a hydrogen bond from donor peptide unit P_i to acceptor peptide unit P_j, we assign the rotational descriptor

$$\mathcal{R}^{H}_{i,j} = (\mathcal{R}_{P_i})^{-1}\mathcal{R}_{P_j} \tag{1}$$

(Fig. 1a and Methods). According to Euler's rotation theorem[9], a rotation is determined by an axis, that is, a vector $\vec{\omega} = (x, y, z)$ of unit length, together with an angle θ (in radians) of rotation around it. We plot such a rotation as a point $\theta\vec{\omega} = (\theta x, \theta y, \theta z)$ in 3D space conveniently imagined as a sphere of radius π representing all rotations, so-called rotational space plots (Fig. 1b, where we stress that antipodal points on the surface sphere of radius π represent the same rotation). This descriptor of a hydrogen bond together with the translation vector rotated by $\mathcal{R}_{\mathrm{P}_i}^{-1}$ (blue arrow, Fig. 1a) from the center of mass of the donor peptide unit to the center of mass of the acceptor peptide unit completely describes the relative positions of a pair of peptide units. However, it is only the rotational part studied here that exhibits the characteristic clustering; in fact, the rotation substantially determines the translation for main chain hydrogen bonds (see discussion below), and hence our 3D rotational descriptor likewise determines the geometry of the hydrogen bond.

Specification of the protein databases. We study three basic classes of databases. Two are derived from PISCES[10] runs with PDB[7] on 12 March 2012:

$$\begin{aligned} \mathrm{HQ} \quad &: \mathrm{Res} \leq 2.0\text{Å and Rfac} \leq 0.2, \\ \mathrm{LQ} \quad &: \mathrm{Res} \leq 3.0\text{Å and Rfac} \leq 0.3, \end{aligned} \tag{2}$$

which are taken at 15, 30, 60 and 95% sequence identity. The third set is the CATH[11] v.4.0.0 library at the levels CATHS, CATHSO, CATHSOL together with 'CATH', a set identified by the CATH developers at the CATH—CATHS level of their database. See Supplementary Note 1 for further specification of the databases. A Dictionary of the Secondary Structure of Proteins (DSSP)[12] hydrogen bond is accepted provided[13] furthermore that:

$$\begin{aligned} &\mathrm{HO} - \mathrm{distance} < 2.7\text{Å}, \\ &\mathrm{angle}\,(\mathrm{NHO}), \mathrm{angle}\,(\mathrm{COH}) > 90^{\circ}. \end{aligned} \tag{3}$$

We remark that our condition for accepting a hydrogen bond depends on the location of the H atom at the amide end in each hydrogen-bonded peptide unit. Please see Supplementary Note 2 for a discussion of the determination of the H atom locations.

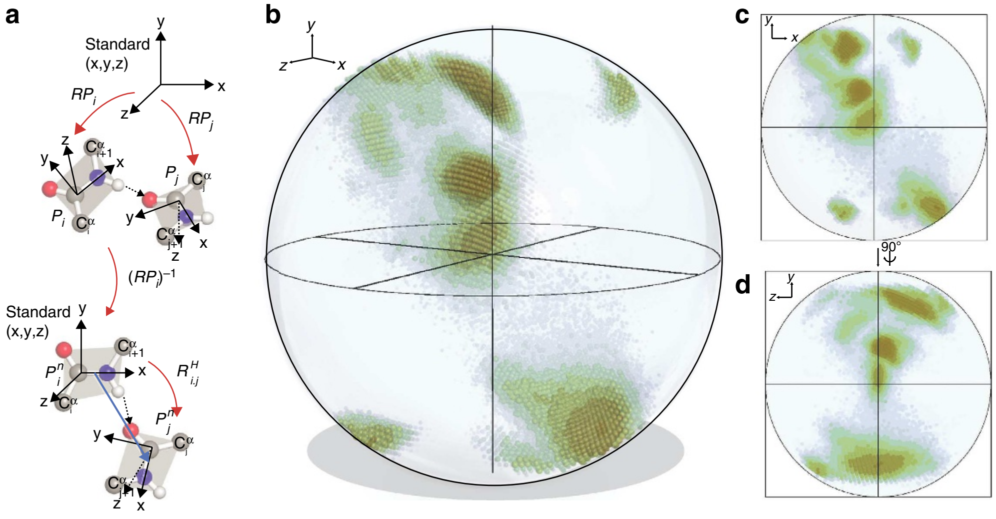

Figure 1 | Density plot of distribution of all hydrogen bond rotations in PDB-HQ60. (**a**) Hydrogen bond from peptide units P_i to peptide unit P_j, where the locations of the backbone atoms C_i^{α}, C_i and N_{i+1} (C_j^{α}, C_j and N_{j+1}) determine a rotation $\mathcal{R}_{P_i}(\mathcal{R}_{P_j})$. Rotation of both peptide units by the inverse of $\mathcal{R}_{P_i}$ brings P_i in standard position, P_i^n and P_j to a new position P_j^n and the rotation bringing P_i^n to P_j^n is the rotation $\mathcal{R}_{i,j}^H = \mathcal{R}_{P_i}^{-1}\mathcal{R}_{P_j}$ associated to the hydrogen bond. The blue arrow from the center of mass of P_i^n to P_j^n is the translation associated to the hydrogen bond. (**b**) 3D rendering of the total distribution coloured by density. (**c**) Orthoscopic projection of the density plot in the *x*-*y* plane. (**d**) Orthoscopic projection of the density plot in the *x*-*z* plane.

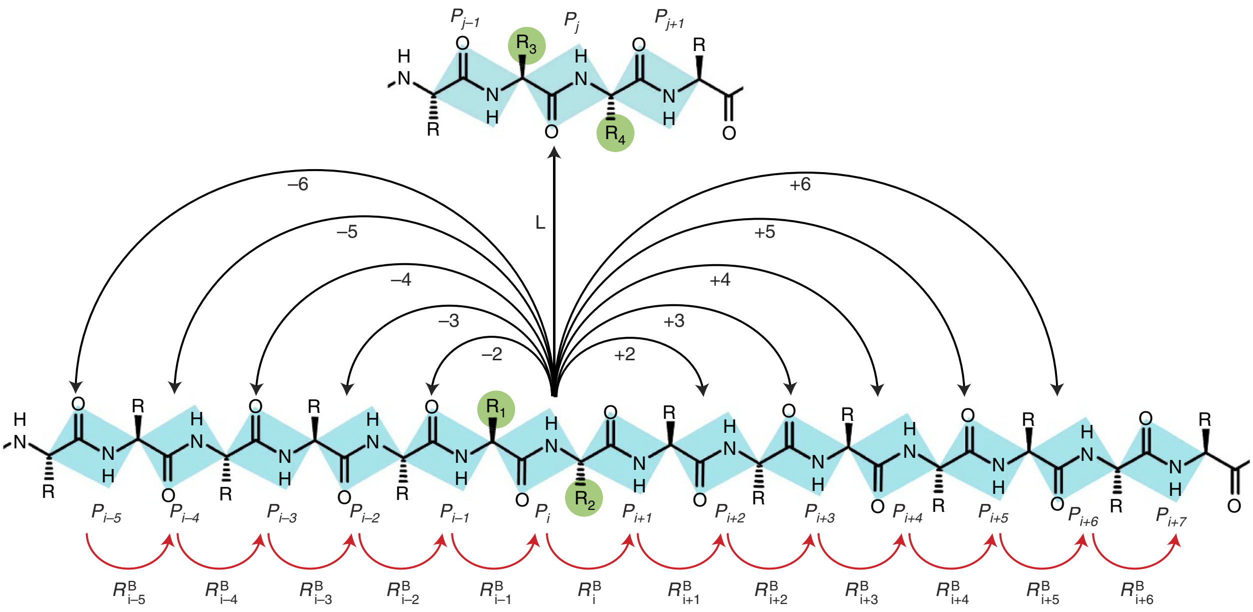

Figure 2 | Indexing conventions of hydrogen bonds. The flanking residues R_1, R_2, R_3 and R_4 around the hydrogen bond from P_i to P_j, the signed length Δ of hydrogen bonds along the backbone and the long-range *L*-group of hydrogen bonds corresponding to $|\Delta| > 6$. The rotations along the backbone $\mathcal{R}_i^B$ indicate the rotations from the *i*th peptide unit P_i to the following P_{i+1}.

Our reference library is HQ60, namely, high-quality data at 60% sequence identity containing 1.16 M hydrogen bonds. A density plot of the raw data (Fig. 1b–d and Supplementary Fig. 1) reveals remarkably that the rotations cluster and concentrate in a relatively small volume of rotational space.

Cluster determination. Several runs of the clustering algorithm detailed under Methods were performed. A first run of the clustering algorithm using all the data together resulted in 14 clusters comprising >99.99% of all the data. Inspection of most of these clusters revealed that the mode box was composed primarily of hydrogen bonds of one fixed length along the backbone. Also, an examination of the translation vectors for these 14 clusters pointed to further limited sub-clustering. For this reason, we refined the analysis by making a separate clustering for each signed difference Δ from donor to acceptor. Precisely, if the hydrogen bond is from donor peptide unit P_i to acceptor peptide unit P_j, then $\Delta = j - i$ if $i < j$ and otherwise $\Delta = j - i - 1$ (Fig. 2).

For $|\Delta| > 6$, we did not find any differences in the clustering results, and thus all data with $|\Delta| > 6$ were combined into a 'long-range' group. Thus, we considered separately $\Delta = -2$, $\Delta = 2$, $\Delta = -3$, $\Delta = 3$, $\Delta = -4$, $\Delta = 4$, $\Delta = -5$, $\Delta = 5$, $\Delta = -6$, $\Delta = 6$ and $|\Delta| > 6$. The clustering algorithm discussed above was run for each of these length categories and resulted in 29 clusters. For each of these clusters, the corresponding translation vectors were considered, and except for the main cluster with $\Delta = -3$, the

Table 1 | Characteristics of the 30 clusters in rotational space based on PDB-HQ60.

Cluster	P(all)	P(mode)	Rotational space Mode	Ave. translation	Volume	Ave. dist.	Max dist.	Peakedness
2_a^-	16,327	203	3.03 (− 0.28, − 0.78, 0.56)	(− 0.73, − 2.27, 1.50)	0.50%	0.268	1.253	0.189
2_b^-	1,249	16	2.86 (0.22, 0.86, 0.46)	(− 0.45, − 2.52, − 1.13)	0.20%	0.285	1.059	0.188
2_c^-	110	3	2.46 (− 0.33, − 0.73, − 0.60)	(− 1.11, − 2.01, − 1.70)	0.03%	0.393	1.283	0.225
3_a^-	164,832	1213	2.50 (− 0.29, 0.92, − 0.28)	(2.05, − 3.66, − 0.27)	1.38%	0.341	1.729	0.240
3_b^-	16,761	156	2.89 (− 0.45, 0.83, − 0.33)	(2.56, − 3.46, 1.46)	0.79%	0.338	1.429	0.234
3_c^-	7,706	77	2.55 (0.31, − 0.87, − 0.38)	(2.03, − 3.76, 0.01)	0.48%	0.292	1.135	0.202
3_d^-	5,490	34	2.86 (0.45, − 0.74, − 0.50)	(2.48, − 3.28, − 1.57)	0.57%	0.416	1.508	0.301
3_e^-	321	6	2.78 (− 0.29, 0.90, 0.32)	(2.09, − 3.13, − 1.97)	0.08%	0.240	0.537	0.185
4_a^-	504,642	16030	1.08 (− 0.32, 0.93, − 0.17)	(2.88, − 3.77, 0.19)	2.42%	0.214	2.045	0.129
4_b^-	1,969	10	2.16 (− 0.84, 0.41, 0.36)	(3.55, − 3.01, − 0.13)	0.68%	1.222	3.139	0.702
5_a^-	16,500	74	2.41 (− 0.63, 0.73, 0.26)	(3.16, − 3.35, − 1.25)	1.21%	0.394	2.797	0.271
5_b^-	3,661	15	0.57 (− 0.64, 0.60, − 0.48)	(3.11, − 3.34, 1.23)	0.67%	0.499	1.967	0.334
5_c^-	3,406	50	2.05 (− 0.56, 0.67, 0.50)	(3.77, − 2.44, − 1.58)	0.29%	0.239	1.26	0.164
5_d^-	1,907	17	0.51 (− 0.62, 0.74, 0.26)	(3.3, − 3.41, − 0.40)	0.30%	0.296	1.218	0.206
5_e^-	295	10	0.77 (− 0.23, 0.96, − 0.18)	(3.01, − 3.57, − 0.52)	0.08%	0.247	1.195	0.147
6_a^-	1,964	7	1.95 (− 0.52, 0.75, 0.40)	(3.11, − 3.07, − 0.89)	0.68%	0.880	3.141	0.430
6_b^-	1,308	11	1.49 (− 0.99, − 0.01, 0.12)	(3.33, − 2.72, 1.56)	0.31%	0.420	1.987	0.235
2_a^+	266	2	2.71 (0.57, − 0.75, − 0.34)	(2.59, − 3.80, − 0.67)	0.12%	0.668	2.001	0.440
3_a^+	6,965	31	2.55 (0.54, − 0.79, − 0.29)	(2.29, − 4.02, − 0.40)	1.10%	0.563	3.139	0.319
3_b^+	2,088	20	2.80 (0.59, − 0.80, 0.08)	(2.63, − 3.81, 0.85)	0.46%	0.452	1.546	0.293
3_c^+	707	5	2.41 (− 0.68, 0.69, − 0.24)	(2.60, − 4.02, 1.00)	0.29%	0.493	1.485	0.331
4_a^+	6,848	24	2.54 (0.52, − 0.83, − 0.21)	(2.38, − 3.92, 0.21)	1.34%	0.605	3.136	0.358
5_a^+	4,525	16	2.67 (0.57, − 0.80, − 0.18)	(2.43, − 3.90, 0.36)	1.10%	0.668	3.120	0.408
6_a^+	1,325	4	2.67 (0.54, − 0.81, − 0.23)	(2.48, − 3.72, 0.69)	0.52%	0.881	2.380	0.647
L_a	242,357	437	2.82 (0.61, − 0.79, 0.00)	(2.44, − 3.90, 0.48)	7.62%	0.591	3.141	0.398
L_b	127,879	227	0.21 (− 0.78, 0.63, 0.06)	(3.00, − 3.51, 0.13)	5.97%	0.591	2.512	0.392
L_c	13,727	33	1.63 (− 0.56, 0.80, − 0.20)	(2.77, − 3.60, − 0.57)	1.81%	0.513	2.322	0.355
L_d	8,747	48	2.79 (− 0.65, 0.70, − 0.30)	(2.35, − 4.04, 1.26)	0.93%	0.383	1.436	0.269
L_e	1,221	7	1.96 (− 0.51, 0.79, 0.35)	(2.84, − 3.25, − 1.52)	0.30%	0.352	1.235	0.272
L_f	808	7	1.84 (− 0.82, 0.35, − 0.45)	(2.79, − 3.21, 2.06)	0.24%	0.370	1.268	0.260

For each cluster, the table lists the total number of points (P(all)), the number of points in the box (see Methods) at the mode in rotational space (P(mode)), the mode point of the rotations (rotational space Mode), average translation (Ave. translation), the volume of the cluster as a percent of the total volume of rotational space (Volume), average (Ave. dist.) and max distance ($_{\text{Max}}$ dist.) from data points to the mode of the cluster and the maximum distance to mode for the top 70% of the density (Peakedness).

translations did not point to any further sub-clustering. For the main cluster with $\Delta = -3$, there was a clear division into two sub-clusters as illustrated in Supplementary Fig. 2. In this way, we ended up with the total of 30 clusters listed in Table 1 and plotted in Fig. 3. The notation for the clusters with $|\Delta| \leq 6$ is

$$|\,\Delta\,|^{\mathrm{Sign}(\Delta)}_{a,b,c,\ldots} \qquad (4)$$

where the subindex is an alpha-numeric enumerator, indexing clusters according to decreasing size (measured in terms of data points in clusters) for a given value of Δ. For the clusters with $|\Delta| > 6$, we use the notation $L_{a,b,c,\ldots}$, where again the alpha-numeric enumerator indexes clusters according to decreasing size.

We have further listed the volume (given in % of the total volume of rotational space), the average and maximal distance within a cluster to the mode and the average distance to the mode among the top 70% of the density. The latter being a measure for how peaked the cluster is with a small value, indicating a peaked distribution and a large value a flat distribution. We observe that even though cluster 4_a^- contains almost half of the hydrogen bonds, it only occupies a little $<$2.5% of the volume. It is the most peaked cluster, about two times more so than 3_a^- and 3.5 times more so than L_a and L_b, which are the other large clusters. We further observe that L_a and L_b have considerably larger volume than any other cluster. We see that 4_b^-, 6_a^- and L_a have the largest possible maximal distance to the mode, namely π. We observe that 93.4% of the data are contained in the seven largest clusters 4_a^-, L_a, L_b, 3_a^-, 3_b^-, 5_a^-, 2_a^-, which together occupy just below 20% of the volume of rotational space. Supplementary Fig. 3 provides for each cluster a plot showing the maximal, minimal and average logarithmic density as a function of distance to the mode. This provides a more detailed description of how peaked the individual clusters are.

In Fig. 4 we provide a sample structure for each of the 30 cluster, whose rotation belongs to the mode of the given cluster.

The original 14 modes from the run with all hydrogen bond rotations together, independent of Δ can be found as the modes of the final clusters 2_a^-, 2_b^-, 2_c^-, 3_a^-, 3_c^-, 3_e^-, 4_a^-, 4_b^-, 5_a^-, 5_c^-, 6_b^-, L_a, L_b and L_d. As can be seen, none of the original modes correspond to the $\Delta > 0$ length categories. This reflects the fact that the modes of $\Delta > 0$ clusters are all close to the modes of the long-range clusters $L_a, \ldots, L_f$.

Clustering has also been analyzed using all of the above-mentioned data sets and the less stringent DSSP[12] definition of hydrogen bonds with substantially the same conclusions as with the high-quality PDB-based libraries in this paper. To further access the stability of the clusters under change of the clustering algorithm, we have also implemented the Mean shift clustering algorithm[14–16], which similarly reproduces the same clustering with minor variations. Please see Supplementary Note 3 for a discussion of the selection of the clustering algorithms and Supplementary Note 4 and Supplementary Table 1 for a comprehensive comparison of all clustering runs.

Hydrogen bonds clusters and Ramachandran plots. To demonstrate that there, in general, are no relations between our 3D rotational descriptor of hydrogen bonds and the Ramachandran plots at the flanking C^{α}, we have in Supplementary Fig. 4 plotted the Ramachandran plots for all the hydrogen bonds in a

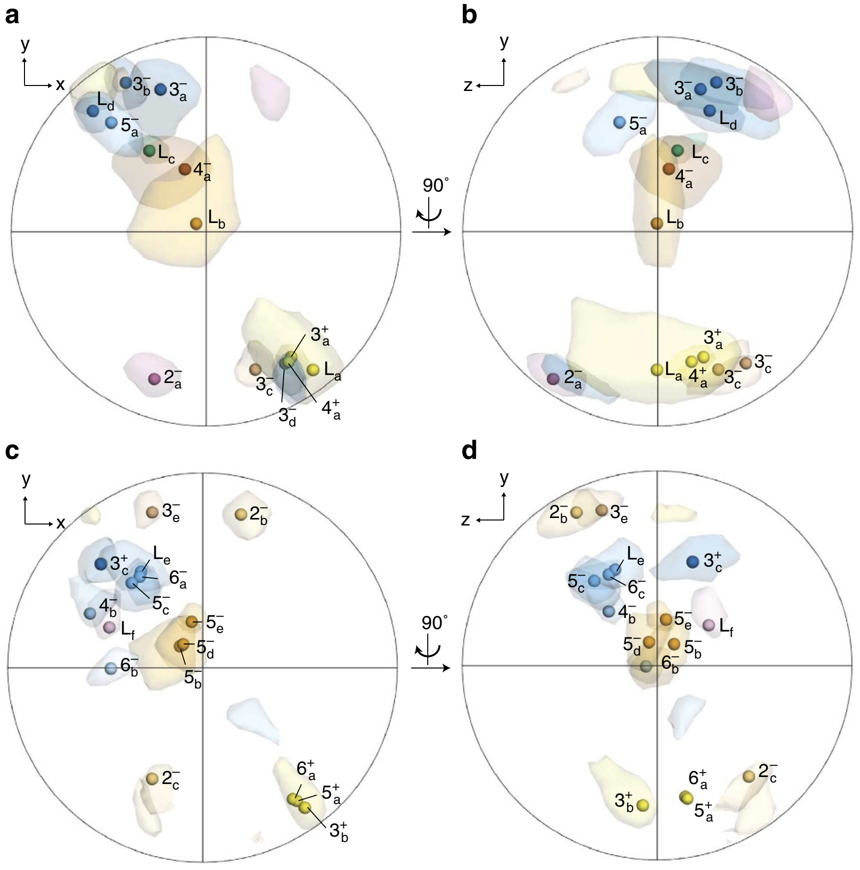

Figure 3 | Plots of the 30 clusters in rotational space. (**a**) and (**b**) are orthoscopic projection plots of the 15 biggest clusters along the *z* and *x* axis. (**c**) and (**d**) are orthoscopic projection plots of the 15 smallest clusters along the *z* and *x* axis.

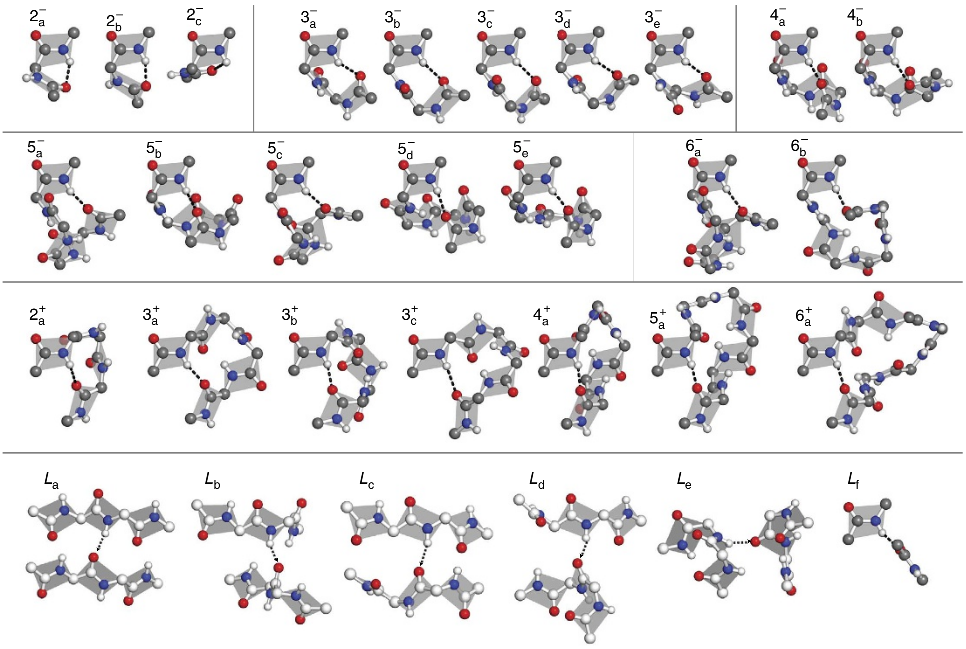

Figure 4 | Sample structure from the mode of each of the 30 clusters. Sample structure from the mode of each of the 30 clusters divided according to length along the backbone.

given cluster for each of the four locations indexed as indicated in Fig. 2.

None of the 15 biggest clusters exhibits a strong localization of any of the four Ramachandran plots. All large clusters and most of the small clusters have a broad spread of the flanking conformational angles. Hence our 3D rotational descriptor of hydrogen bonds is akin to the Ramachandran plots but is by no means determined by the flanking Ramachandran plots.

Table 2 | Distances between cluster modes and idealized gamma/beta/alpha/pi-turns.

Mode	Gamma +	Gamma −
2_a^-	2.30	0.19
2_b^-	0.21	2.09
2_c^-	0.90	2.42

Mode	3_{10}	I	I′	II	II′	VIa1	VIA2	VIb	VIII
3_a^-	0.19	0.38	1.89	0.95	1.72	1.91	1.28	1.79	2.20
3_b^-	0.67	0.41	1.75	0.46	1.66	1.84	1.00	1.84	2.28
3_c^-	1.88	2.02	0.18	1.88	0.80	0.68	1.16	2.87	2.41
3_d^-	1.99	2.10	0.39	2.02	0.35	0.17	1.01	2.67	2.23
3_e^-	1.27	1.52	1.00	1.76	0.71	0.83	1.05	2.99	2.89

Mode	Alpha helix (R)	Alpha helix (L)	IRS	ILS	IIRS	IILS	IRU	ILU	IIRU	IILU
4_a^-	0.01	2.64	0.43	1.90	0.69	2.64	1.98	2.89	2.70	2.43
4_b^-	1.67	2.91	1.34	2.98	1.62	2.59	0.47	2.02	1.36	1.76

Mode	Pi Helix	HB-AAAa	HB-PgAA	HB-AAAA	Schellman
5_a^-	2.22	0.37	1.93	1.79	0.19
5_b^-	0.44	2.37	0.65	1.09	2.04
5_c^-	1.88	0.89	1.67	1.38	0.48
5_d^-	0.32	2.27	0.65	0.84	1.89
5_e^-	0.60	2.17	0.99	1.26	1.87

Certain clusters do, however, have significantly higher density at certain conformational angles. The large cluster 3_a^- has a peak at the conformational angles of the 3_{10} helix, 4_a^- at the α helix and L_a and L_b at the β-strand conformational angles. We also in part see this for some of the short-range clusters at certain of the positions $R_1,\ldots,R_4$, for example, for clusters 2_b^-, 2_c^-, 3_e^-, 5_d^-, 5_e^-, 6_b^-. It is, however, still clear that we do not even in these cases see a localized single cluster behaviour, except for the following very small clusters 5_d^-, 5_e^-, 3_b^- (R_4) and 3_c^- (R_3). All of this fits well with the following analysis of the primary and secondary imprints of our clusters and the following correlation analysis between our clusters compared with known structural motives.

Characterizing primary/secondary structure imprints. To explore the primary/secondary propensities of the clusters, we evaluate the primary and secondary structure of the four residues of the donor and acceptor peptide units. The residues are annotated R_1, R_2, R_3 and R_4 as in Fig. 2, where R_2 contains the donor amide and R_3 contains the donor carbonyl. The residue distribution for the clusters is plotted in the left column of Supplementary Fig. 5.

Clusters 2_b^- and 2_c^- have a 30% occurrence of Glycine at position $R_1 = R_4$. 3_b^- and 3_c^- have a 70% occurrence of Glycine at position R_1, whereas 3_d^- has a 65% occurrence of Glycine at R_4. The cluster 3_e^- has 95% Proline at R_1. These findings correspond closely to the known residue preferences for γ and β turns. Cluster 4_b^- has a 20–40% occurrence of Glycine at $R_1,\ldots,R_4$ showing that this longer turn motif requires a Glycine to bend the backbone, with the position of the Glycine being less important. As we will observe below, 5_a^-, 5_c^- and 6_a^- have similar hydrogen bond rotational geometry, which is also reflected here by a shared preference for Glycine at R_1 of 60–95% and a preference for Leucine at R_2 of 20–35%. A few smaller preferences are observed for the +-clusters: 2_a^+ has a preference for Glycine at R_1 of 40% and R_2 of 20%; 3_c^+ has 25% Serine at R_2; 4_a^+ and 5_a^+ have a 20–30% preference for Aspartic acid at position R_2. Long-range clusters L_c, L_e and L_f show a 20–40% preference at R_1 or R_2. In conclusion, the primary sequence of the clusters is only a weak signal characterized by Glycine, Proline and other amino acids that provide special backbone conformations.

The primary sequence signatures of residue four-tuples were likewise investigated, but no highly occurring four-tuple patterns were observed (please see Supplementary Table 2).

Next, we investigate the secondary structure patterns of the clusters by using the DSSP[12] annotation. DSSP annotates residues with seven secondary structure classes based on backbone geometry: H = α helix, B = residue in isolated beta-bridge, E = extended strand, participating in beta ladder, G = 3-helix (3_{10} helix), I = 5-helix (π helix), T = hydrogen bonded turn, S = bend, – = unclassified. The secondary structure preferences were plotted in Supplementary Fig. 5 (right column). Cluster 2_a^- is identified as unclassified by DSSP, whereas 2_b^- and 2_c^- are identified as S or T, respectively. Cluster 3_a^- shows a signal for 3_{10} helix, α helix and turn on R_{1-4}, which is in accordance with our ideal correlation of both 3_{10} helix and beta-turn type I discussed below. Clusters 3_b^- and 3_e^- have the pattern T–T, whereas 3_c^- and 3_d^- are predisposed to TEET and correlated to beta turns. Cluster 4_a^- has a very strong HHHH signature, which corresponds to α helices consistent with the further analysis below. Cluster 4_b^- has a weaker DSSP predisposition, but shows preference of 40% for H at R_{124} and 50% for T at R_{12}. All 5^- and 6^- clusters have a strong preference for HH at R_{34}. Cluster 2_a^+ has ESSE pattern, whereas the remaining +-cluster and L-cluster have mainly E (beta) annotations.

Turns and helixes. For hydrogen bonds between peptide units in helixes and turns, we can compute the corresponding rotations as described under Methods. The natural distance measure between two rotations is the geodesic distance also recalled in the Method section. Table 2 lists the distances between the modes of clusters with backbone length Δ between -2 and -5 and the respective ideal structures (Fig. 5). Thus, the modes of cluster 2_a^- and 2_b^- closely correspond to the ideal gamma turns (Fig. 5a). Cluster 4_a^- has its mode exactly at the ideal α helix, whereas the mode of cluster 4_b^- is close to the alpha turn $\mathrm{I}-\alpha_{\mathrm{RU}}$ (Fig. 5c). Alpha turn IRS lies in 4_a^-, whereas the remaining alpha turns lie in 4_b^-. For

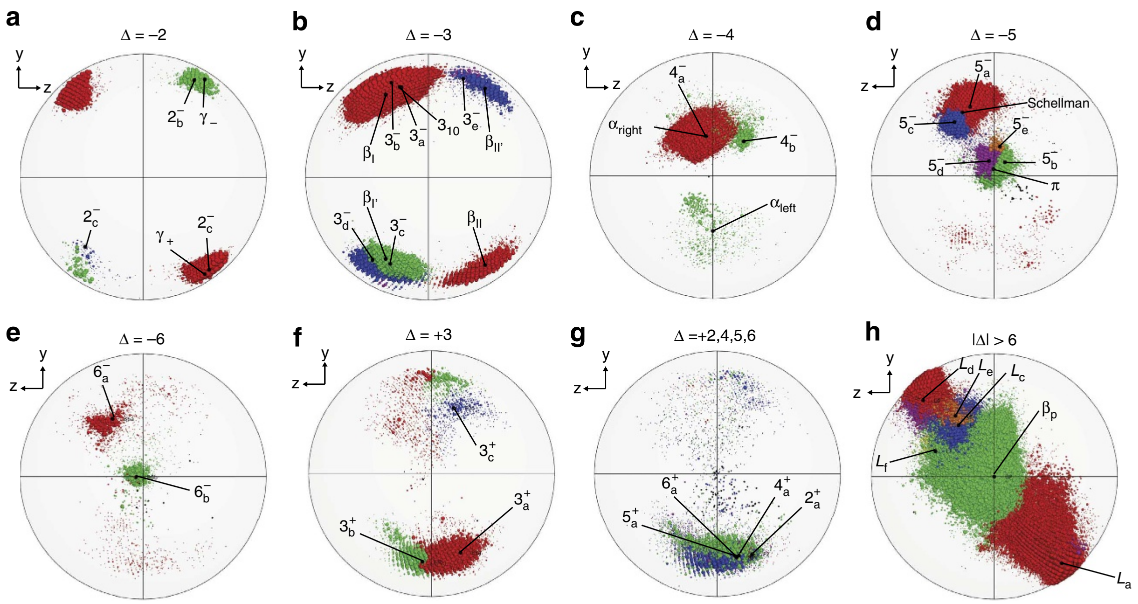

Figure 5 | Comparison between ideal motives and clusters. For each signed length Δ along the backbone, the clusters are plotted (in red, green, blue and orange) with indication of the mode location of the cluster together with the location of some of the ideal structures in rotational space: $\gamma\pm$ for 2^-, β_I, β_{II}, $\beta_{I'}$, $\beta_{II'}$ for 3^-, α_{right}, α_{left} for 4^-, Schellman, π for 5^- and β_p for L. The letters **a** to **h** indicate the different length categories.

Table 3 | Relation on mode point rotations for long-range clusters and extended β-strands.

Relation	Orientation	L_a	L_b	L_c	L_d	L_e	L_f
$\mathcal{R}_a^B\mathcal{R}^{-1}\mathcal{R}_a^B$	Anti	0.38	0.57	0.30	0.42	0.76	1.33
$(\mathcal{R}_a^B)^{-1}\mathcal{R}^{-1}(\mathcal{R}_a^B)^{-1}$	Anti	0.52	0.28	0.29	0.28	0.28	0.80
$\mathcal{R}_p^B\mathcal{R}^{-1}\left(\mathcal{R}_p^B\right)^{-1}$	Parallel	0.43	0.11	0.45	0.45	0.49	0.69
$\mathcal{R}_a^B\mathcal{R}^{-1}\mathcal{R}_p^B$	Anti	0.30	0.63	0.28	0.23	0.73	1.11
$\left(\mathcal{R}_p^B\right)^{-1}\mathcal{R}^{-1}(\mathcal{R}_a^B)^{-1}$	Anti	0.40	0.43	0.52	0.16	0.35	0.60
$(\mathcal{R}_a^B)^{-1}\mathcal{R}^{-1}\mathcal{R}_p^B$	Parallel	0.52	0.35	0.53	0.50	0.66	0.98
$\left(\mathcal{R}_p^B\right)^{-1}\mathcal{R}^{-1}\mathcal{R}_a^B$	Parallel	0.57	0.30	0.29	0.42	0.59	0.85
Rotation angle		2.82	0.21	1.63	2.79	1.96	1.84

The table list the distances between the modes $\mathcal{R}$ and the transformed mode as given by one of the relations. Included are also the rotation angles (in radians) for the mode point of the clusters.

length $\Delta = -5$ clusters, the mode of 5_a^- is particularly close to the Schellman motif and also close to HB-AAAa. The modes of clusters 5_b^-, 5_d^-, 5_e^- are closets to the ideal π helix rotation and 5_c^- is closets to the Schellman motif. Pi turn HB-PgAA lies in cluster 5_b^- and HB-AAAA in 5_d^- (Fig. 5d). The cluster to which the exact ideal hydrogen bond rotation either belongs or is closest to, for each motif, is given in the last column of Supplementary Table 3.

To further investigate the beta turns, we have for each of the five clusters $3_a^-,\ldots,3_e^-$ calculated the occurrences of the eight turn types (Supplementary Table 4). Using the conformational angles in Supplementary Table 3, a turn belongs to a turn type[17] if three angles deviate $<30°$ from the ideal values and the last deviates by at most 45°.

From the Supplementary Table 4, it appears that cluster 3_a^- is dominated by 3_{10} helices and beta turn I, cluster 3_b^- by beta turn II, 3_c^- by beta turn I′, 3_d^- by beta turn II′, and cluster 3_e^- by beta turn VIa1 although this turn is found in 3_d^- and 3_a^- as well. The distances from the mode points of the clusters to the ideal structures confirm these findings (Fig. 5b).

To estimate the proportion of the bonds in clusters 3_a^-, 4_a^- and 5_b^-, 5_d^-, 5_e^-, which are involved in helical patterns, we have computed the proportion of the bonds in the cluster, which are flanked by bonds from the same length category. As it appears from Supplementary Table 5, 80% (of which 17.5% are at ends) of the bonds in 4_a^- are involved in α-helixes, 39.5% (of which 35.5% are at ends) in 3_{10} helices. The majority of bonds from 5_d^- and half of the bonds in 5_e^- and ∼30% of the bonds from 5_b^- are involved in π-helices.

Beta turns VIa2, VIb and VIII are special in the sense that the hydrogen bonds calculated from the backbone conformational angles assuming idealized geometry are not to be found in the data set of hydrogen bonds. A closer study reveals that indeed these turns do not conform to idealized geometry, and hence one can simply not compute their rotations using conformational angles. The rotational space distance between the hydrogen bond rotation and the rotation calculated from the backbone conformational angles as $\mathcal{R}^H = (\mathcal{R}_T^B(2))^{-1}(\mathcal{R}_T^B(1))^{-1}$ assuming idealized geometry (and cis conformation in peptide plane two for VIa2 and VIb) are of the order 1–3 in distance instead of the order 0.2 in distance when idealized geometry holds. This demonstrates further the utility of over rotational descriptor, which in this case is much more accurate than backbone conformational angles.

Table 4 | Probing for Beta hairpins in the various clusters.

Cluster	Total	Hairpin	Candidate hydrogen bonds $\mathcal{R}$	Position of $\mathcal{R}_1$	Distance
3_a^-	164,823	2:2	479	3_c^+	0.26
3_b^-	16,759	2:2	366	$3_b^+,3_c^+$	0.77, 0.26
3_c^-	7,706	2:2	1495	3_a^+	0.30
3_d^-	5,489	2:2	195	3_a^+	0.51
4_a^-	504,563	3:3	143	None	
4_b^-	1,969	3:3	80	4_a^+	0.79
5_a^-	16,498	4:4	131	5_a^+	0.63
6_a^-	1,963	5:5	127	6_a^+	0.77
3_a^+	6,965	2:4	661	L_a	0.55
3_b^+	2,088	2:4	774	L_a	0.47
3_c^+	707	2:4	31	L_d	0.43
4_a^+	6,848	3:5	750	L_a	0.41
5_a^+	4,525	4:6	655	L_a	0.40
6_a^+	1,325	5:7	169	L_a	0.43

Total number of bonds in the cluster (Total), type of hairpin (Hairpin), number of bonds with conformational angles around the ideal antiparallel (Candidate hydrogen bonds), cluster identification for the next hydrogen bond in an antiparallel beta strand, where the next hydrogen bond is calculated from either (15) or (16) (Position of $\mathcal{R}_1$), and rotational space distance of the next hydrogen bond to the mode of the cluster. Only clusters with at least 10 candidate hydrogen bonds have been included.

Long-range clusters. For long-range hydrogen bonds, we can of course not proceed in the same way as for turns and helixes, moving very long distances along the backbone. We proceed instead inductively as described under methods.

Using the notation from the Method section, Table 3 list the distances between $\mathcal{R}^H$ and $\mathcal{R}_1^H$ with $\mathcal{R}^H$, one of the six mode points for the long-range clusters and $\mathcal{R}_1^H$ is given by either (15), (16) or (17) (with the distance the same for the two parts of the latter). The table also considers the relations appropriate for the case of a parallel and antiparallel beta strand being part of a beta sheet. The relations then involve both of $\mathcal{R}_a^B$ and $\mathcal{R}_p^B$.

The mode of cluster L_b is close to the ideal extended parallel beta strand. The modes of clusters L_a and L_d are both close to the ideal extended antiparallel case, and both have a rotation angle close to π. Clusters L_c, L_e and L_f all have an intermediate rotation angle. The mode of cluster L_c is close to solving both of (15) and (16), whereas the mode of cluster L_e is close to solving the latter only (Fig. 5h). The mode of cluster L_f seems unrelated to extended beta strands. Actually, L_f is the only cluster for which the transformed mode $\mathcal{R}_1$ of the mode point $\mathcal{R}$ is closer to the mode point of another cluster than to $\mathcal{R}$ itself for the two antiparallel transformations (being closer to the mode point of L_c, the distances being 1.11 and 0.77 from the two first entries of Table 3). The mode of cluster L_d is very close to solving the equations when a parallel and antiparallel beta strand are neighbours in an antiparallel fashion. The mode of cluster L_a has the same property, whereas the mode of cluster L_b fits the corresponding equations when the two strands are neighbours in a parallel fashion.

Looking at the Ramachandran plots for all of the clusters in Supplementary Fig. 4, we see that, apart from the extended parallel and antiparallel beta strands, each cluster contains hydrogen bonds not related to these structures. Please see Supplementary Fig. 6 and Supplementary Note 5 for further discussion of this point.

Beta hairpins. Hydrogen bonds from beta hairpins will by their very nature appear in several clusters. Take as an example a 2:2 hairpin, where the inner hyrogen bond of the loop is a Beta turn (3^- clusters) and the next hydrogen bond along the strand belongs to one of the 3^+ clusters.

Table 4 gives the number of potential beta hairpins of the form described under Methods in the various clusters and the cluster annotation of the next hydrogen bond $\mathcal{R}_1^H$ along the beta strand. Among 2:2 hairpins, as defined above, more than half are found in the combination $\left(3_c^-, 3_a^+\right)$ with the second largest group in the combination $\left(3_a^-, 3_c^+\right)$. There are a few cases of 4:4 and 5:5 hairpins giving rise to the combinations $\left(5_a^-, 5_a^+\right)$ and $\left(6_a^-, 6_a^+\right)$. Hairpins involving the long-range clusters are primarily found in L_a, namely in the combinations $\left(3_a^+, L_a\right)$, $\left(3_b^+, L_a\right)$, $\left(4_a^+, L_a\right)$, $\left(5_a^+, L_a\right)$ and $\left(6_a^+, L_a\right)$. The only exception is a small number of 2:4 hairpins in the combination $\left(3_c^+, L_d\right)$. The existence of beta hairpins in 3_a^+, 4_a^+, 5_a^+, 6_a^+ helps explain the proximity of the modes of these clusters and the mode of long-range cluster L_a.

Accessible part of the rotational space. To assess, from a steric viewpoint, which part of rotational space is available for hydrogen bonding, we performed the following computation: all atoms of two free peptide units were described as rigid spheres and a search of translations was made to identify if the hydrogen bond criterion from DSSP and condition (3) could be met without any overlap of the rigid spheres occurring for non-covalently bonded atoms.

For each grid point (grid size 2π/81) inside the sphere of radius π, up to 21,952 different translations were tested against the hydrogen bond recognition criteria. If a translation resulting in an acceptable hydrogen bond was found, then the above test was performed. With these constraints, it was found that 95% of rotational space had at least one possible translation that resulted in a hydrogen bond (Fig. 6c). The observed high-density regions in rotational space are thus not a consequence of steric constraints alone. As the default restrained refinement of structures obtained by X-ray crystallography only includes a standard set of stereo-chemical restraints (covalent bonds, angles, dihedrals, planarities, chiralities, non-bonded), the just mentioned analysis further demonstrates that the observed clustering cannot be seen as a consequence of this refinement process either.

Local hydrogen bond energy landscapes. To probe the (non)-locality of the formation of the clusters in rotational space, we modelled hydrogen bonds between backbone peptide units by Density Function Theory (DFT), which has proven to be successful in describing basic secondary structure motifs[18,19]. We first probe the energy landscape of rotational space by modelling two peptide units as described under Methods.

The resulting energy landscape (Fig. 6a), which in this case is twofold symmetric, describes, to some extend, the same overall part of rotational space as the experimental PDB based clustering shown in (Fig. 1b–d). Because of symmetry, the global minimum appears twice in rotational space. Two bonding classes are defined by which the O-lone pair is used (the methyl-side or the nitrogen side of the carbonyl) (Fig. 6a). The two classes of hydrogen bonding between two *N*-methylacetamide molecules span two large volumes of rotational space, meaning that the bonding is rather insensitive to the relative rotation of the two molecules; intrinsic properties of the hydrogen bond define the overall volume of spatial rotations. However, we must conclude that the fine clustering, which our analysis of PDB results in, does not arise from two free isolated peptide units, interacting in a single hydrogen bond. This strongly highlights the non-locality and importance of this observed clustering of rotations across hydrogen bonds in protein structures.

Next we analyzed the influence of local backbone constraints on the energy landscape by modelling two fused methylacetamide molecules, which further interacts via an hydrogen bond from the

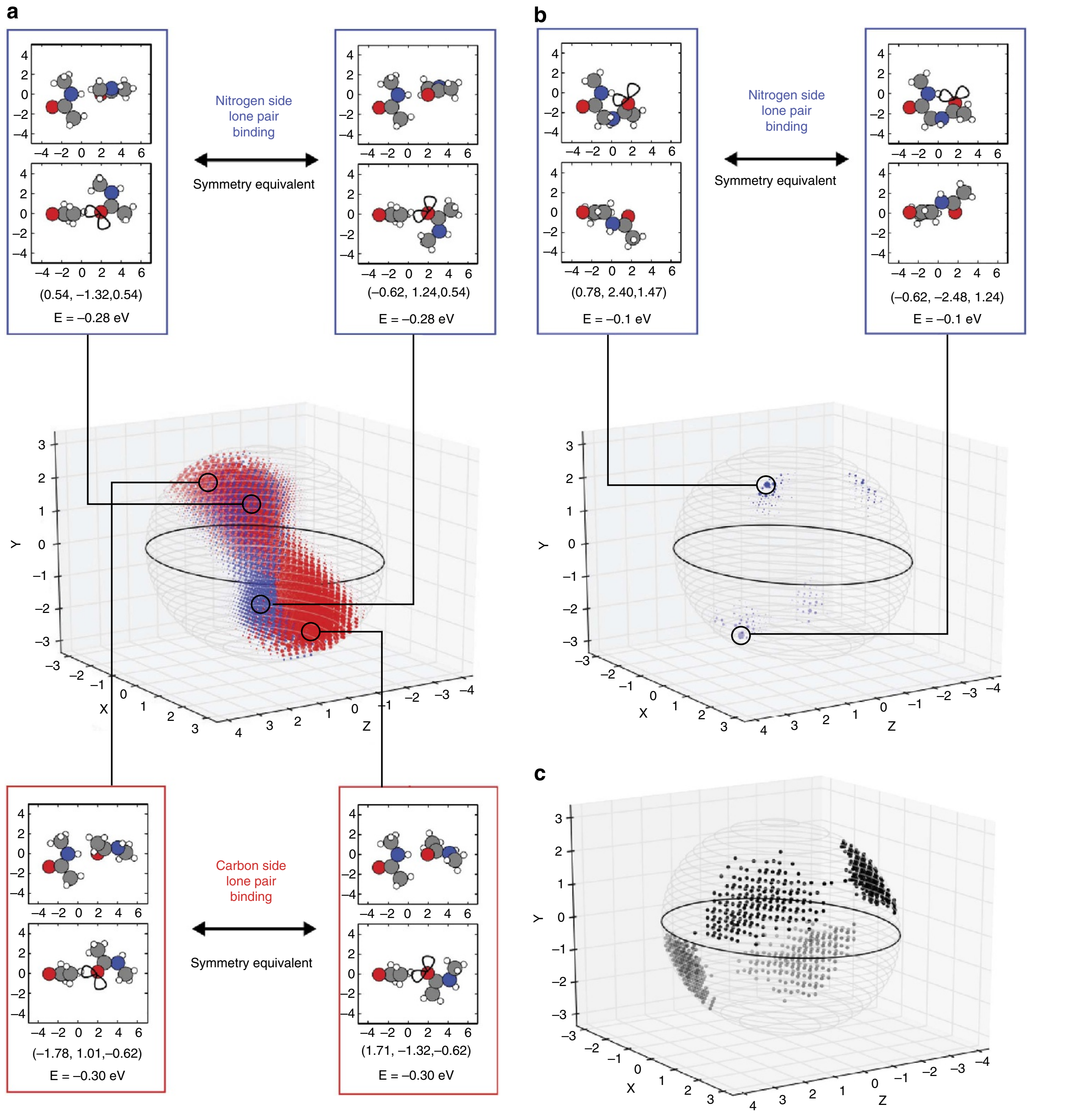

Figure 6 | DFT calculations for CH3-NH-CO--CH3 dimer and CH3-NH-CO-CH2-NH-CO-CH3. (**a**) For each sampled point in rotational space, a sphere is plotted whose radius is proportional to the Boltzmann weight factor exp(− E/kT) of that point at room temperature with the normalization constant chosen such that points with 0.1 eV above the global minimum of − 0.30 eV vanish. By this procedure, hydrogen bonds of at least 0.2 eV strength appear coloured in the figure and are found to comprise 32% of the entire volume of rotational space. Assessing the volume taken up by hydrogen bonds of at least 0.1 eV strength, one reaches 50% of the volume of rotational space.The most stable bonding class is that of the acceptor molecule binding through the O-lone pair, which is at the methyl side of the carbonyl (red spheres in the figure). The other significant bonding class is that of the acceptor molecule binding through the O-lone pair, which is at the amide side of the carbonyl (blue spheres in the figure). Both are doubly degenerate because of symmetry. (**b**) This plot is prepared just as (**a**) is, but here for two consecutive peptide units along the backbone, and only one bonding class (modulo symmetry) is identified. These two minima are recognized as γ_+ and γ_- turns, respectively. (**c**) Plot of the volume in rotational space (black points) with prohibitive steric overlap of rigid atomic spheres. The transparent volume in the plot represents the 95% of rotational space where hydrogen bonding is allowed for steric reasons.

one peptide unit to the immediate next one along the backbone (only possible with this short backbone). The configurations were relaxed as discussed under Methods, but the spanning of rotational space was limited to situations in which the length of the hydrogen bond was initially smaller than 3.5 Å; this excludes fully extended structures, which were also not included in our analysis of the PDB. The resulting energy landscape (Fig. 6b) shows two minima corresponding to γ_+ and γ_- turns, which correspond to cluster 2_b^- and 2_a^-, respectively (Fig. 5a). The energy minima are now relatively weaker, which demonstrates the compromise between covalent backbone bonding and hydrogen bonds, where the high force constants of the covalent bonds present between the two C–C–N units drastically reduce the accessible configurations. This result together with the finding of a widespread minimum of the hydrogen bond in general (Fig. 6a) is strongly suggestive that the clusters empirically observed in protein structures are in part caused by geometric implications of the polypeptide backbone structure of the protein together with the identity of the residues, that is, the protein primary structure.

Discussion

Our rotational descriptor of main chain hydrogen bonds has been demonstrated to be an effective coordinate on the possible local geometries around hydrogen bonds and the geodesic distance in rotation space a useful measure of the discrepancy between two hydrogen bonds. The descriptor provides a uniform viewpoint on local configurations of peptide units participating in an hydrogen bond encompassing secondary structures and turns, and it displays a remarkable clustering over all hydrogen bonds in the PDB. Akin to Ramachandran plots, 3D plots in rotational space usefully depict the relative positions of peptide units across hydrogen bonds relevant for studying protein conformation, dynamics and pathways. Our overall hydrogen bond patterns could be used to annotate protein secondary and tertiary structure, which may lead to a valuable and robust new classification of protein folds[11,20].

NMR is a widely used technique for determining structures of proteins in solution[21] where it is common practice to calculate an ensemble of structures. The scatter within this ensemble derives from both genuine dynamics of the protein in solution and from lack of experimental constraints. Traditionally, this scatter is viewed by superimposing the structures and measuring coordinate differences[22]. However, this can be cumbersome and even misleading if large unstructured regions are present that mask more important structural features, for example, a dynamical hinge connecting two rigid regions. Our classification is a powerful tool for analyzing structural ensembles derived by X-ray/NMR, as visualization of rotations highlights important structural features while unstructured regions are omitted, for example, our method pinpoints only the dynamical part of a hinge.

X-ray crystallography can yield very accurate structures, which are modelled to fit the observed electron-density maps. These maps, that integrate all available diffraction data, may contain less ordered local regions, and the quality of derived structures is overall limited by the resolution of the diffraction data and the accuracy of the determined phases. Among a number of validation tools and procedures to ensure model quality is the unbiased Ramachandran plot. Entirely analogously, our new constraints on hydrogen bond rotations provide hydrogen bond plot quality as a new tool to have a complementary role in model validation and refinement, especially in cases where phi–psi angle restraints defer any use of Ramachandran plot validation such as low-resolution crystallography, electron microscopy and *in silico* modelling.

The further geometrically possible beta strands, we find here, comprised of clusters L_c, L_e and L_f perhaps pose an interesting opportunity in *de novo* protein-structure design.

As demonstrated[23] for RNA, free energies coupled to the topology of fatgraphs[8] labelled by nucleic acids can be effectively used to predict RNA secondary structure. Hydrogen bond free energy[5,24] relying on the distribution of hydrogen bond rotations within each cluster could be readily implemented as Boltzmann statistics based on HQ60 for example. Coupled to chord diagrams with chords labelled by cluster and the backbone labelled by amino acids, this could provide a new tool for *ab initio* protein folding.

As rotations can be assigned to any ordered pair of peptide units, relationships between them beyond hydrogen bonding, such as spatial proximity, can be likewise studied. Suitable triples of vectors can moreover be similarly assigned to any oriented covalent bond and rotations used to study relationships between them. Our basic method could therefore be much more broadly applied to include protein side-chains or general ligands for example.

A web-based implementation of our descriptor for uploaded PDB files is anonymously available at `http://bion-server.au.dk/hbonds/`.

Methods

Peptide plane rotation. Associate a triple $\mathcal{F}_{P_i} = (\vec{u}_i, \vec{v}_i, \vec{w}_i)$ of three-dimensional vectors to the peptide unit P_i of a protein containing the consecutive backbone atoms $\mathrm{C}_i^\alpha - \mathrm{C}_i = \mathrm{N}_{i+1} - \mathrm{C}_{i+1}^\alpha$ in the usual crystallographic notation as follows:

$$\vec{u}_i = \frac{\vec{z}_i - \vec{y}_i}{|\vec{z}_i - \vec{y}_i|}, \ \vec{v}_i = \frac{\vec{y}_i - \vec{x}_i - (\vec{u}_i \cdot (\vec{y}_i - \vec{x}_i))\vec{u}_i}{|\vec{y}_i - \vec{x}_i - (\vec{u}_i \cdot (\vec{y}_i - \vec{x}_i))\vec{u}_i|}, \ \vec{w}_i = \vec{u}_i \times \vec{v}_i \tag{5}$$

In standard vector notation, where $\vec{x}_i$, $\vec{y}_i$, $\vec{z}_i$ are the respective coordinates of C_i^α, C_i,N_{i+1}. Such a triple $\mathcal{F}_{P_i} = (\vec{u}_i, \vec{v}_i, \vec{w}_i)$ is described by a 3-by-3 matrix $\mathcal{R}_{P_i}$ the respective columns of which are the coordinates of $\vec{u}_i$, $\vec{v}_i$, $\vec{w}_i$ in the standard vector basis. For any two peptide units P_i and P_j, with corresponding matrices $\mathcal{R}_{P_i}$ and $\mathcal{R}_{P_j}$, the rotation $\mathcal{R}_{P_j}(\mathcal{R}_{P_i})^{-1}$ brings $\mathcal{F}_{P_i}$ to $\mathcal{F}_{P_j}$. However, this rotation will change if we rotate the entire protein. A descriptor, which is independent of overall rotation is obtained on transforming both of $\mathcal{F}_{P_i}$ and $\mathcal{F}_{P_j}$ by $\mathcal{R}_{P_i}^{-1}$. Thus, $\mathcal{F}_{P_i}$ becomes the standard vector basis and $\mathcal{F}_{P_j}$ becomes $\mathcal{R}_{i,j} = \mathcal{R}_{P_i}^{-1}\mathcal{R}_{P_j}$. We use $\mathcal{R}_{i,j}^H = \mathcal{R}_{i,j}$ as our descriptor for the rotation bringing P_i to P_j. For three peptide units P_i, P_j, P_k, we have the relation $\mathcal{R}_{i,k} = \mathcal{R}_{i,j}\mathcal{R}_{j,k}$.

In the special case when the two peptide units $P_{\mathrm{i}-1}$ and P_i are consecutive along the backbone sharing the carbon C_i^α, the rotation matrix $\mathcal{R}_{i-1}^B = \mathcal{R}_{i-1,i}$ is a function of the backbone conformational angle φ_i preceding and ψ_i following C_i^α. Assuming the idealized geometry of exact tetrahedral angles among bonds at each alpha carbon atom and 120-degree angle between bonds within a peptide unit, one finds[8],

$$\mathcal{R}_{i-1}^B = \tilde{\mathcal{R}}^B(\varphi_i)\bar{\mathcal{R}}^B(\varphi_i + \psi_i)\begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & 0 & -1 \end{pmatrix}, \tag{6}$$

where, with $C_1 = \cos\varphi$ and $S_1 = \sin\varphi$,

$$\tilde{\mathcal{R}}^B(\varphi) = \begin{pmatrix} \frac{2}{3} - \frac{C_1^2}{3} + \frac{S_1^2}{6} & -2\left[\frac{\sqrt{2}C_1}{3} + \frac{S_1^2}{4\sqrt{3}}\right] & 2\left[\frac{C_1 S_1}{2\sqrt{3}} - \frac{S_1}{3\sqrt{2}}\right] \\ 2\left[\frac{\sqrt{2}C_1}{3} - \frac{S_1^2}{4\sqrt{3}}\right] & \frac{2}{3} - \frac{C_1^2}{3} - \frac{S_1^2}{6} & -2\left[\frac{C_1 S_1}{6} + \frac{S_1}{\sqrt{6}}\right] \\ 2\left[\frac{C_1 S_1}{2\sqrt{3}} + \frac{S_1}{3\sqrt{2}}\right] & 2\left[\frac{S_1}{\sqrt{6}} - \frac{C_1 S_1}{6}\right] & \frac{2}{3} + \frac{C_1^2}{3} - \frac{S_1^2}{3} \end{pmatrix} \tag{7}$$

and, with $C_2 = \cos\frac{\varphi+\psi}{2}$ and $S_2 = \sin\frac{\varphi+\psi}{2}$,

$$\tilde{\mathcal{R}}^B(\varphi + \psi) = \begin{pmatrix} 1 - \frac{3}{2}S_2^2 & \frac{\sqrt{3}}{2}S_2^2 & \sqrt{3}C_2S_2 \\ \frac{\sqrt{3}}{2}S_2^2 & 1 - \frac{1}{2}S_2^2 & -C_2S_2 \\ -\sqrt{3}C_2S_2 & C_2S_2 & 1 - 2S_2^2 \end{pmatrix}. \tag{8}$$

When peptide unit P_{i-1} is in *cis* conformation, the matrix $\mathcal{R}_{i-1}^B$ must be premultiplied by a diagonal matrix with entries $(1, -1, -1)$. Supplementary Table 3 lists the conformational angles defining the secondary structures considered in this paper from which the idealized backbone transition $\mathcal{R}_{i-1}^B$ is calculated.

Rotations and angle-axis pairs. Let $\vec{e}_1 = (1,0,0)$, $\vec{e}_2 = (0,1,0)$, $\vec{e}_3 = (0,0,1)$ be the usual vector basis of space. A rotation must necessarily map these vectors to another respective triple $\vec{u}_1, \vec{u}_2, \vec{u}_3$ of pairwise perpendicular vectors of length one with cross-product $\vec{u}_1 \times \vec{u}_2 = \vec{u}_3$. We may write $\vec{u}_r = \sum_{s=1}^3 a_{sr}\vec{e}_s$, for each $r = 1, 2, 3$, to determine the matrix $\mathcal{R} = (a_{rs})$, where the respective columns are the coordinates of $\vec{u}_1, \vec{u}_2, \vec{u}_3$. The trace of $\mathcal{R}$ is the sum tr $\mathcal{R} = a_{11} + a_{22} + a_{33}$ of its diagonal entries.

One can compute the matrix $\mathcal{R} = \mathcal{R}(\theta, \vec{\omega})$ corresponding to a given angle-axis pair $\theta, \vec{\omega}$ as follows. First, associate to $\vec{\omega} = (u, v, w)$ the matrix Ω with rows $(0, \ -w, v))$, $(w, 0, \ -u)$ and $(-v, u, 0)$, and finally define $\mathcal{R} = I + \Omega \sin\theta + \Omega^2(1 - \cos\theta)$, where I denotes the 3-by-3 identity matrix, and $\Omega^2 = \Omega\Omega$ denotes the matrix product. Conversely, the angle-axis pair $\theta, \vec{\omega}$ corresponding to the matrix $\mathcal{R}$ is given by $\theta = \arccos(\frac{\mathrm{tr}\mathcal{R}-1}{2})$, $0 \leq \theta \leq \pi$ with $\vec{\omega}$, the vector of unit length proportional to $(a_{32}-a_{23}, a_{13}-a_{31}, a_{21}-a_{12})$ for the case $0<\theta<\pi$. When $\theta=0$, the matrix A is the identity matrix and $\vec{\omega}$ is undetermined. When $\theta=\pi$, the axis $\vec{\omega}$ is determined up to its sign only by the equations $u^2-1=a_{11}$, $v^2-1=a_{22}$, $w^2-1=a_{33}$, $2uv=a_{12}$, $2uw=a_{13}$ and $2vw=a_{23}$.

To describe distances between the modes of the clusters and known secondary structures, it is useful to define the transpose to be the matrix $\mathcal{R}^t = (a_{sr})$ derived from $\mathcal{R} = (a_{rs})$ by interchanging rows and columns. The geodesic distance on rotational space between two rotations $\mathcal{R}, \mathcal{R}'$ is determined by the angle in the angle-axis pair associated to the rotation $\mathcal{R}(\mathcal{R}')^t$, that is, the distance between $\mathcal{R}$ and $\mathcal{R}'$ is

$$d(\mathcal{R}, \mathcal{R}') = \left| \arccos\left(\frac{\mathrm{tr}(\mathcal{R}, \mathcal{R}')^t - 1}{2} \right) \right|. \tag{9}$$

Hence distances in rotational space are between 0 and π. A special property is bi-invariance in the sense that $d(\mathcal{R}_1\mathcal{R}\mathcal{R}_2, \mathcal{R}_1\mathcal{R}'\mathcal{R}_2) = d(\mathcal{R}, \mathcal{R}')$ for any four rotations $\mathcal{R}, \mathcal{R}', \mathcal{R}_1, \mathcal{R}_2$, that is, the distance is unchanged under matrix multiplication on both the right and left sides. Furthermore, one can show that the bi-invariant (un-normalized Haar) measure on the collection of rotations is given by $\frac{2(1-\cos r)}{r^2} dxdydz$, where $r = \sqrt{x^2+y^2+z^2}$ in the coordinates (x, y, z) of our plots with total volume of rotational space equal to $8\pi^2$. Thus, densities of points in our plots must be appropriately scaled, and distances between rotations must be computed using $d(\mathcal{R}, \mathcal{R}')$. In particular, this means that a true evenly distribution on rotation space will in our presentation also result in an evenly distributed density.

The clustering algorithm. To perform the clustering (see Supplementary Note 3 for choice of clustering algorithm) of the rotations, it is convenient to transform all data points in rotational space by left multiplication with an element $\mathcal{R}_0^{-1}$ in rotational space to move most of the observed rotations away from the boundary sphere. The convenient rotation $\mathcal{R}_0$ is given by 2.479, (−0.282, 0.907, −0.313) in angle-axis notation, which is a point of density with angle coordinate fairly close to π. Next, the cube $(-\pi, \pi)^3$ is divided into 81 × 81 × 81 small 'boxes' with side lengths $\delta = 2\pi/81$, so box (n, m, p) has center (x_n, y_m, z_p) with $x_n = -\pi + (n-0.5)\delta$ and y_m and z_p defined similarly. Let n_{nmp} denote the number of transformed data points within the box (n, m, p) if it lies entirely within the ball B of radius π. For a box B at the boundary of the sphere of radius π, we count the number of points in B as well as the neighbouring antipodal box. The density in a box is given by $d_{nmp} = \frac{\theta_{nmp}^2}{2(1-\cos(\theta_{nmp}))} n_{nmp}$, where $\theta_{nmp} = \sqrt{x_n^2+y_m^2+z_p^2}$.

To form the clusters, the algorithm first finds seeds for the clusters after the boxes have been ordered according to density. A box with a local maximum of the density becomes a seed if the P-value for testing equal rates in boxes with distance 1 and boxes with distance 2, to the box under consideration, is below 0.003, and the P-value for the similar test using boxes with distances 2 and 3 is below 0.01 (for the robustness of these conditions please see Supplementary Fig. 7). To make the procedure robust to the number of boxes used to divide $(-\pi, \pi)^3$ and to enhance the possibility of finding clusters in low-density regions, a similar run is made with a 64 × 64 × 64 division. If two nearby seeds from the first run are joined only one is kept as a final seed, and if a new seed is found, new meaning that it is sufficiently apart from the seeds of the first round, this is included as a seed in the 81 × 81 × 81 division (see Supplementary Note 4 for a comparison of clusterings for varying box sizes). The method used corresponds to testing for a known probability of success in a binomial distribution. A large P-value indicates that there is not much difference between boxes with distance 1 and boxes with distance 2 (or boxes with distance 2 and boxes with distance 3), that is, the density is fairly flat, not pointing to a well-defined cluster mode, hence the test. A small P-value on the other hand points to a cluster with a well-defined mode point. Having established the seeds, the densities d_{nmp} for varying n, m, p are ordered according to decreasing size. The algorithm adds one box at a time, where each added box becomes a member of a cluster if the box distance from the point to the cluster is one or two; if there are several competing clusters to join, we choose the one with the highest density for the box closest to the new point. Having run through all nonempty boxes, a second run is made to allow unclassified boxes from the first run to join nearby clusters when the rotational space distance is <1.5 times the width of a box.

Except for the minor clusters 2_c^-, 2_a^+, 3_e^-, 3_c^+, 5_b^-, $5_{e_-}^-$, 6_a^+, L_e and L_d, both of the P-values used to define a seed of a cluster are $<10^{-5}$ (Supplementary Fig. 7).

Hydrogen bond rotations in turns and helixes. The hydrogen bond rotation between peptide units i and j can be calculated from the transformations $\mathcal{R}_k^B = (\mathcal{R}_{P_k})^{-1}\mathcal{R}_{P_{k+1}}$ along the backbone. Specifically (Fig. 2), if the donor peptide unit is after the acceptor unit along the backbone, then $i>j$ and the hydrogen bond rotation is

$$\mathcal{R}_{i,j}^H = (\mathcal{R}_{i-1}^B)^{-1}(\mathcal{R}_{i-2}^B)^{-1} \cdots (\mathcal{R}_j^B)^{-1}; \tag{10}$$

if the donor peptide unit is before the acceptor unit along the backbone, then $i<j$ and the hydrogen bond rotation is

$$\mathcal{R}_{i,j}^H = \mathcal{R}_i^B \ldots \mathcal{R}_{j-1}^B. \tag{11}$$

To correlate various clusters in rotational space with known structural motives, we use the rotations given by the formulae (6), (7) and (8) for $\mathcal{R}_k^B$ as functions of conformational angles and the known conformational angles for various local loops, turns and helix structures (Supplementary Table 3). For gamma turns, the hydrogen bond rotations are

$$\mathcal{R}^H = (\mathcal{R}_{\gamma+}^B)^{-1} \text{ and } \mathcal{R}^H = (\mathcal{R}_{\gamma-}^B)^{-1} \tag{12}$$

respectively, where $\mathcal{R}_{\gamma+}^B$ and $\mathcal{R}_{\gamma-}^B$ are the backbone transformations defined from the set of conformational angles listed in Supplementary Table 3. For beta turns, the hydrogen bonds rotations are

$$\mathcal{R}^H = \mathcal{R}_T^B(2)^{-1}\mathcal{R}_T^B(1)^{-1}, \tag{13}$$

where T is the turn type, and the backbone rotations $\mathcal{R}_T^B(\cdot)$ are derived from two sets of conformational angles listed in Supplementary Table 3. For 3_{10}, α and π helices, the hydrogen bond rotations are

$$\mathcal{R}^H = (\mathcal{R}_{3_{10}}^B)^{-2}, (\mathcal{R}_\alpha^B)^{-3} \text{ and } (\mathcal{R}_\pi^B)^{-4} \tag{14}$$

respectively, where $\mathcal{R}_{3_{10}}^B$, $\mathcal{R}_\alpha^B$ and $\mathcal{R}_\pi^B$ are the respective backbone rotations. Supplementary Table 3 gives a complete list of the hydrogen bond rotations for all the considered secondary structures and turns assuming idealized geometry and exact *cis* or *trans* conformation.

Hydrogen bond rotations in long-range clusters. For the long-range clusters, we consider their association with ideal parallel or antiparallel beta strands. For an antiparallel beta strand, if there is an hydrogen bond between residue i and j with rotation $\mathcal{R}^H$, the next hydrogen bond further away from the loop has a rotation $\mathcal{R}_1^H$ given by either

$$\mathcal{R}_1^H = \mathcal{R}_a^B(\mathcal{R}^H)^{-1}\mathcal{R}_a^B, \tag{15}$$

or

$$\mathcal{R}_1^H = (\mathcal{R}_a^B)^{-1}(\mathcal{R}^H)^{-1}(\mathcal{R}_a^B)^{-1}, \tag{16}$$

again under the assumption of ideal conformational angles along the backbone. For an extended antiparallel beta strand, we require $\mathcal{R}_1^H = \mathcal{R}^H$. The corresponding relations for an ideal parallel beta strand are

$$\mathcal{R}_1^H = \mathcal{R}_p^B(\mathcal{R}^H)^{-1}\left(\mathcal{R}_p^B\right)^{-1} \quad \text{and} \quad \mathcal{R}_1^H = \left(\mathcal{R}_p^B\right)^{-1}(\mathcal{R}^H)^{-1}\mathcal{R}_p^B \tag{17}$$

where $\mathcal{R}_p^B$ is the ideal transformation along the backbone given in Supplementary Table 6. See Supplementary Fig. 8 for an explanation of these relations. For the extended parallel case, we require $\mathcal{R}_1^H = \mathcal{R}^H$. If $\mathcal{R}^H = \mathcal{R}_p^B(\mathcal{R}^H)^{-1}(\mathcal{R}_p^B)^{-1}$, then we also have $\mathcal{R}^H = (\mathcal{R}_p^B)^{-1}\mathcal{R}^{-1}\mathcal{R}_p^B$, and the only solutions are $\mathcal{R}^H = \mathrm{Id}$ and a half–full turn around the axis of $\mathcal{R}_p^B$, the latter not being relevant for a parallel beta strand.

If the conformational angles between the two hydrogen bonds correspond to an ideal antiparallel beta strand, then we can calculate the rotation $\mathcal{R}_1^H$ of the next hydrogen bond from the rotation $\mathcal{R}^H$ of the inner hydrogen bond by equation (15), where $\mathcal{R}_a^B$ is the ideal transformation along the backbone. Next, consider the situation where the inner hydrogen bond of the loop belongs to one of the + clusters (2:4 hairpin). In this case, the next hydrogen bond belongs to the long-range clusters. Still assuming that the conformational angles between the two hydrogen bonds correspond to an ideal antiparallel beta strand, we get equation (16) instead of equation (15). We have probed for the existence of pairs $\mathcal{R}^H$ and $\mathcal{R}_1^H$ subject to either (15) or (16) as follows. For each turn cluster, we find candidate $\mathcal{R}^H$'s as those hydrogen bonds with conformational angles outwards of the turn close to the ideal antiparallel conformational angles (−135°, 150°) allowing for a deviation of 30°. From the selected cases, we calculate an average hydrogen bond rotation $\mathcal{R}^H$, find $\mathcal{R}_1^H$ from the above equations (15) and (16), and look for this rotation in the appropriate length category of hydrogen bonds.

Density function theory. DFT has been used to investigate the nature of the hydrogen bonds between backbone peptide units. The calculations were done using the ASE/GPAW package[25] using projector augmented waves and a real space basis (periodic boundary conditions in a 19.2 × 19.2 × 19.2 Å super cell and a grid-spacing of 0.16 Å). Exchange-correlation effects were described using the Perdew-Burke-Ernzerhof (PBE) functional. This functional has been proven successful in describing the hydrogen binding within, for example, helical polypeptides (including the transitions from the alpha-helix to the pi- and 3–10 helices)[18] and in describing the side-group propensities within beta-sheets[19].

We probe the entire energy landscape of rotational space by modelling two peptide units, minimalistically represented by methylacetamide, CH_3–NH–CO–CH_3. The relative position of the donor and acceptor is calculated for each position of rotational space followed by relaxation of all atomic degrees of freedom (except for the C–C–N coordinates, whose relative position are fixed and only allowed to translate as a rigid unit, so as to not change the rotation from one unit to the other).

References

1. Pauling, L. *The Nature of the Chemical Bond* (Cornell University Press, 1960).
2. Fersht, A. *Enzyme Structure and Function* (Freeman, 1985).
3. Bordo, D. & Argos, P. The role of side-chain hydrogen bonds in the formation and stabilization of secondary structure in soluble proteins. *J. Mol. Biol.* **243,** 504–519 (1994).
4. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7,** 95–99 (1963).
5. Morozov, A. V., Kortemme, T., Tsemekhman, K. & Baker, D. Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc Natl Acad Sci. USA* **101,** 6946–6951 (2004).
6. Grishaev, A. & Bax, A. An empirical backbone-backbone hydrogen-bonding potential in proteins and its applications to NMR structure refinement and validation. *J. Am. Chem. Soc.* **126,** 7281–7292 (2004).
7. Protein Data Bank, http://www.rcsb.org/pdb/.
8. Penner, R. C., Knudsen, M., Wiuf, C. & Andersen, J. E. Fatgraph models of proteins. *Comm. Pure Appl. Math* **63,** 1249–1297 (2010).
9. Euler, L. Formulae generales pro translatione quacunque corporum rigidorum. *Novi Commentarii academiae scientiarum Petropolitanae* **20,** 189–207 (1776).
10. Wang, G. & Dunbrack, Jr. R. L. PISCES: a protein sequence culling server. *Bioinformatics* **19,** 1589–1591 (2003).
11. Orengo, C. A. *et al.* CATH-a hierarchic classification of protein domain structures. *Structure* **5,** 1093–1108 (1997).
12. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 (1983).
13. Baker, E. N. & Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **44,** 97–179 (1984).
14. Chen, Y. Mean shift, mode seeking, and clustering. *IEEE. Trans. Pattern. Anal. Mach. Intell.* **17,** 790–799 (1995).
15. Li, J., Ray, S. & Lindsay, B. G. A nonparametric statistical approach to clustering via mode identification. *J. Mach. Learn. Res.* **8,** 1687–1723 (2007).
16. Meer, P & Subbarao, R. Nonlinear mean shift over Riemannian manifolds. *Int. J. Comput. Vis.* **84,** 1–20 (2009).
17. Chou, K. C. Prediction of tight turns and their types in proteins. *Anal. Biochem.* **286,** 1–16 (2000).
18. Ireta, J., Neugebauer, J., Scheffler, M., Rojo, A. & Galvan, M. Structural transitions in the polyalanine alpha-helix under uniaxial strain. *J. Am. Chem. Soc.* **127,** 17241–17244 (2005).
19. Rossmeisl, J., Kristensen, I., Gregersen, M., Jacobsen, K. W. & Norskov, J. K. *β*-sheet preferences from first principles. *J. Am. Chem. Soc.* **125,** 16383–16386 (2003).
20. Andreeva, A. *et al.* Data growth and its impact on the SCOP database: new developments. *Nucleic. Acids. Res.* **36,** D419–D425 (2008).
21. Wuthrich, K. (ed.) *NMR of proteins and nucleic acids* (Wiley, 1986).
22. O'Donoghue, S. I. *et al.* Visualization of Macromolecular Structures. *Nat. Methods.* **7,** S42–S55 (2010).
23. Reidys, C. M. *et al.* Topology and prediction of RNA pseudoknots. *Bioinformatics* **27,** 1076–1085 (2011).
24. Morozov, A. V. & Kortemme, T. Potential functions for hydrogen bonds in protein structure prediction and design. *Adv. Prot. Chem.* **72,** 1–38 (2006).
25. Enkovaara *et al.* Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *J. Phys. Condens. Matter* **22,** 253202 (2010).

Acknowledgements

All researchers in this project were supported by the Danish National Research Foundation through the foundations centre of excellence grants 'Centre for Quantum Geometry of Moduli Spaces', 'Centre for DNA nanotechnology', 'Centre for Membrane Pumps in Cells and Disease' and 'Center for Insoluble Protein Structures' from Danish National Research Foundation. R.C.P. and J.E.A. acknowledge valuable discussions with Alexey Finkelstein during the early stages of this work. We thank Niels Larsen for technical assistance in setting up a web analysis server, and Rikke Schmidt-Kjaergaard for graphical suggestions and assistance and Rasmus Villemoes for assistance with programming.

Author contributions

The abstract mathematical model was conceived jointly by R.C.P. and J.E.A. Initial programming on the project was conducted by R.R. and the initial indication of the overall clustering was observed. E.S.A., A.K.K., M.B., N.C.N., P.N. and J.T.N. was added to the project, to define the used databases and to design the precise extract from the PDB databases together with the specification of chemical correct hydrogen bonds. Precise correlation of PDB-files with DSSP files was programmed by J.T.N., who also programmed parsers, which computed all needed extracts from these databases, based on R.R.'s initial code. J.L.J. was attached to the project to conduct all statistical analysis and he programmed the clustering algorithms. R.C.P., E.S.A., J.L.J., A.K.K., M.B., P.N., N.C.N. and J.E.A. have contributed significantly in the data analysis as far as evaluation of hydrogen bonds in concrete examples, study of correlation between clusters and primary, secondary and turn structures and the further analysis of the geometry of hydrogen bonds contained in each individual cluster. The DFT analysis was conducted by A.M.H.R., K.L.S. and B.H. with input from E.S.A. and J.E.A.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: R.C.P. and J.E.A. share personal financial interests on US patent applications related to applying moduli space techniques to the analysis of biomolecules.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Penner, R.C. *et al.* Hydrogen bond rotations as a uniform structural tool for analyzing protein architecture. *Nat. Commun.* 5:5803 doi: 10.1038/ncomms6803 (2014).

nature COMMUNICATIONS

DOI: 10.1038/ncomms7209

Erratum: Hydrogen bond rotations as a uniform structural tool for analyzing protein architecture

Robert C. Penner, Ebbe S. Andersen, Jens L. Jensen, Adriana K. Kantcheva, Maike Bublitz, Poul Nissen, Anton M.H. Rasmussen, Katrine L. Svane, Bjørk Hammer, Reza Rezazadegan, Niels Chr Nielsen, Jakob T. Nielsen & Jørgen E. Andersen

Nature Communications 5:5803 doi: 10.1038/ncomms6803 (2015); Published 17 Dec 2014; Updated 3 Feb 2015

The authors Robert C. Penner and Ebbe S. Andersen were incorrectly omitted from the list of corresponding authors in this Article. The correct information for correspondence is: 'Correspondence and requests for materials should be addressed to R.C.P. (email: rpenner@qgm.au.dk) or to E.S.A. (email: esa@inano.au.dk) or to J.E.A. (email: andersen@qgm.au.dk)'. The Article also contains errors in the labelling of Figs 1 and 2. In Fig. 1a, the labels RP_i, RP_j, $(RP_i)^{-1}$ and $R^H_{i,j}$ should read $\mathcal{R}_{P_i}$, $\mathcal{R}_{P_j}$, $(\mathcal{R}_{P_i})^{-1}$ and $\mathcal{R}^H_{i,j}$, respectively. In Fig. 2, labels R^B_{i-5}, R^B_{i-4}, R^B_{i-3}, R^B_{i-2}, R^B_{i-1}, R^B_i, R^B_{i+1}, R^B_{i+2}, R^B_{i+3}, R^B_{i+4}, R^B_{i+5} and R^B_{i+6} should read $\mathcal{R}^B_{i-5}$, $\mathcal{R}^B_{i-4}$, $\mathcal{R}^B_{i-3}$, $\mathcal{R}^B_{i-2}$, $\mathcal{R}^B_{i-1}$, $\mathcal{R}^B_i$, $\mathcal{R}^B_{i+1}$, $\mathcal{R}^B_{i+2}$, $\mathcal{R}^B_{i+3}$, $\mathcal{R}^B_{i+4}$, $\mathcal{R}^B_{i+5}$ and $\mathcal{R}^B_{i+6}$, respectively. The correct versions of these figures follow.

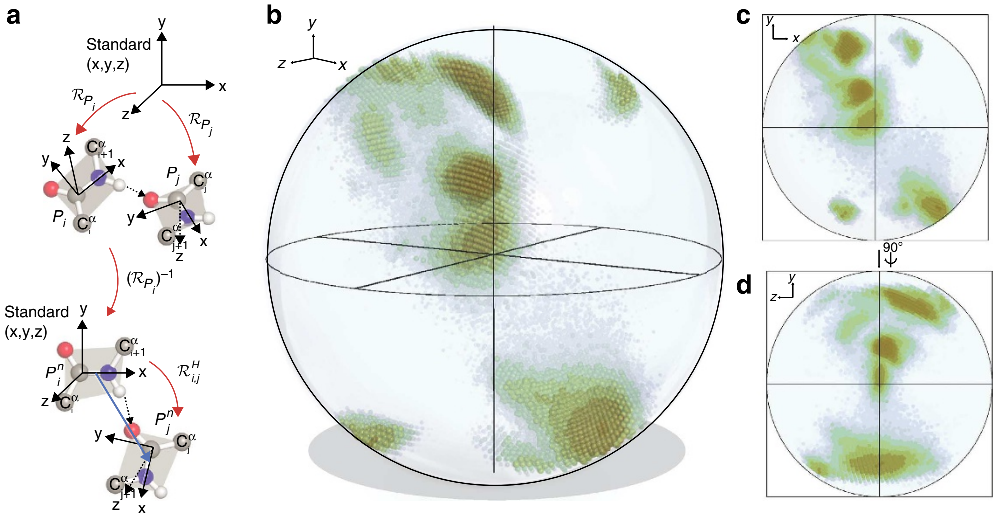

Figure 1

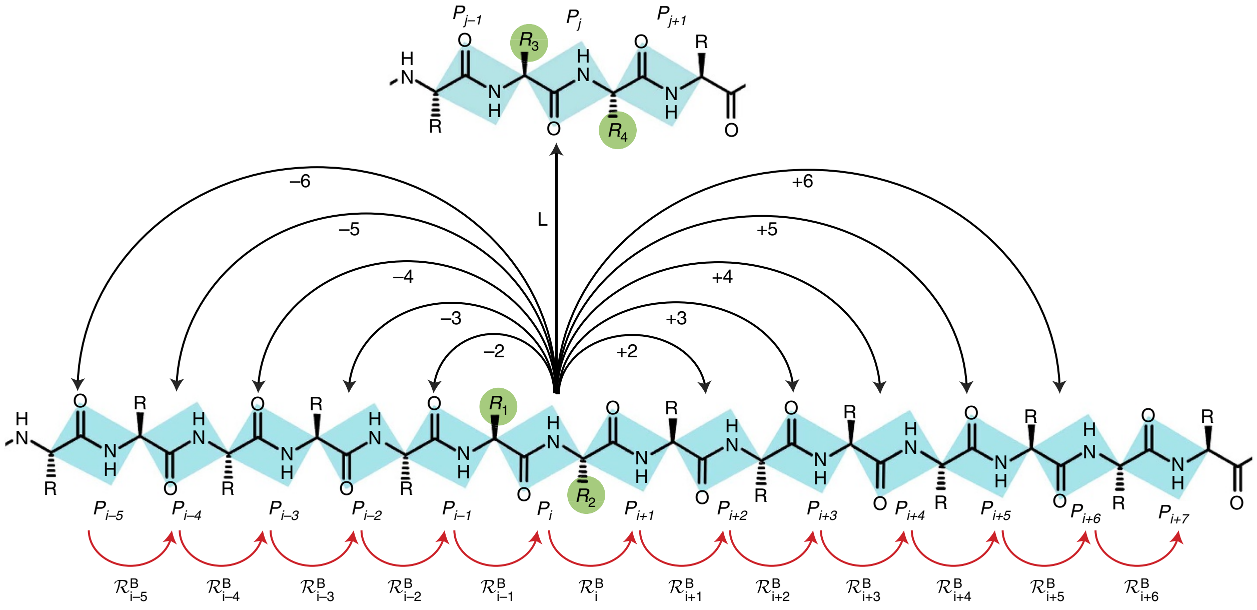
P_{j-1}
P_j
P_{j+1}
R_3
R_4
L
−6
−5
−4
−3
−2
+2
+3
+4
+5
+6
R_1
R_2
P_{i-5}
P_{i-4}
P_{i-3}
P_{i-2}
P_{i-1}
P_i
P_{i+1}
P_{i+2}
P_{i+3}
P_{i+4}
P_{i+5}
P_{i+6}
P_{i+7}
$\mathcal{R}^B_{i-5}$
$\mathcal{R}^B_{i-4}$
$\mathcal{R}^B_{i-3}$
$\mathcal{R}^B_{i-2}$
$\mathcal{R}^B_{i-1}$
$\mathcal{R}^B_{i}$
$\mathcal{R}^B_{i+1}$
$\mathcal{R}^B_{i+2}$
$\mathcal{R}^B_{i+3}$
$\mathcal{R}^B_{i+4}$
$\mathcal{R}^B_{i+5}$
$\mathcal{R}^B_{i+6}$

Figure 2